Bernoulli **18**(1), 2012, 279–289
DOI: 10.3150/10-BEJ337

On the inclusion probabilities in some unequal probability sampling plans without replacement

YAMING YU

Department of Statistics, University of California, Irvine, CA 92697, USA. E-mail: yamingy@uci.edu

Comparison results are obtained for the inclusion probabilities in some unequal probability sampling plans without replacement. For either successive sampling or Hájek's rejective sampling, the larger the sample size, the more uniform the inclusion probabilities in the sense of majorization. In particular, the inclusion probabilities are more uniform than the drawing probabilities. For the same sample size, and given the same set of drawing probabilities, the inclusion probabilities are more uniform for rejective sampling than for successive sampling. This last result confirms a conjecture of Hájek (*Sampling from a Finite Population* (1981) Dekker). Results are also presented in terms of the Kullback–Leibler divergence, showing that the inclusion probabilities for successive sampling are more proportional to the drawing probabilities.

Keywords: conditional Poisson sampling; entropy; Hájek's conjecture; sampling without replacement; stochastic orders; total positivity order

1. Introduction and main results

Consider a finite population indexed by $U = \{1, \dots, N\}$. Let $\alpha = (\alpha_1, \dots, \alpha_N)$, $\sum_{i=1}^N \alpha_i = 1$, denote a set of drawing probabilities. In Hájek's [5,6] *rejective sampling*, independent draws are made with probabilities according to the same α until a sample of size n is obtained; whenever a duplicate appears, all draws are rejected and the process restarts. *Successive sampling*, a closely related scheme, makes the same independent draws except that whenever a duplicate appears, only the current draw is rejected and needs to be redrawn. Mathematically, rejective sampling is equivalent to conditional Poisson sampling, that is, independent sampling for each unit conditional on the sample size being n. Conditional Poisson sampling possesses a maximum entropy property, among other desirable properties, and has received considerable attention; see Chen, Dempster and Liu [3], Berger [2], Traat, Bondesson and Meister [23], Arratia, Goldstein and Langholz [1], and Qualité [15]. It also has interesting applications to modeling how players select lottery tickets [22]. Successive sampling, on the other hand, has connections to areas such as software reliability [11].

Unequal probability sampling may achieve considerable variance reduction if the first-order inclusion probabilities are made proportional to a suitable auxiliary variable. For either rejective sampling or successive sampling, however, the inclusion probabilities are rather complicated and generally not proportional to the drawing probabilities α. Thus relationships between the inclusion probabilities and α, either approximate or exact, are of interest. This work considers exact qualitative comparisons. See Hájek [5] and Rosén [18–20] for asymptotic results.

1350-7265

Denote the inclusion probabilities for rejective sampling by $\pi^{\mathrm{R}} = (\pi_1^{\mathrm{R}}, \ldots, \pi_N^{\mathrm{R}})$ and those for successive sampling by $\pi^{\mathrm{S}} = (\pi_1^{\mathrm{S}}, \ldots, \pi_N^{\mathrm{S}})$. Hájek [6], page 97, conjectures the following inequalities based on asymptotic considerations and numerical experience:

$$\frac{\max \alpha_i}{\min \alpha_i} \geq \frac{\max \pi_i^{\mathrm{S}}}{\min \pi_i^{\mathrm{S}}} \geq \frac{\max \pi_i^{\mathrm{R}}}{\min \pi_i^{\mathrm{R}}}.$$

Milbrodt [14] proposes a strengthened conjecture,

$$n \max \alpha_i \geq \max \pi_i^{\mathrm{S}} \geq \max \pi_i^{\mathrm{R}}, \tag{1}$$

$$n \min \alpha_i \leq \min \pi_i^{\mathrm{S}} \leq \min \pi_i^{\mathrm{R}}, \tag{2}$$

and partially resolves it by showing

$$\begin{aligned} &n \max \alpha_i \geq \max \pi_i^{\mathrm{S}}, \qquad n \min \alpha_i \leq \min \pi_i^{\mathrm{S}}, \\ &n \max \alpha_i \geq \max \pi_i^{\mathrm{R}}, \qquad n \min \alpha_i \leq \min \pi_i^{\mathrm{R}}. \end{aligned} \tag{3}$$

The inequalities (3) are also obtained by Rao, Sengupta and Sinha [16]. The inequalities $\max \pi_i^{\mathrm{S}} \geq \max \pi_i^{\mathrm{R}}$ and $\min \pi_i^{\mathrm{S}} \leq \min \pi_i^{\mathrm{R}}$ have remained open; see Milbrodt [14] for numerical illustrations. Roughly speaking, both Hájek's conjecture and Milbrodt's strengthened version say that the drawing probabilities are more variable than the inclusion probabilities for successive sampling, which are themselves more variable than the inclusion probabilities for rejective sampling.

Concerning successive sampling, Kochar and Korwar [12] obtain some comparison results using the notion of majorization. A real vector $b = (b_1, \ldots, b_N)$ is said to majorize $a = (a_1, \ldots, a_N)$, written as $a \prec b$, if

- $\sum_{i=1}^N a_i = \sum_{i=1}^N b_i$, and
- $\sum_{i=k}^N a_{(i)} \leq \sum_{i=k}^N b_{(i)}, k = 2, \ldots, N,$ where $a_{(1)} \leq \cdots \leq a_{(N)}$ and $b_{(1)} \leq \cdots \leq b_{(N)}$ are $(a_1, \ldots, a_N)$ and $(b_1, \ldots, b_N)$ arranged in increasing order, respectively.

Kochar and Korwar [12] show that

$$n^{-1} \pi^{\mathrm{S}} \prec \alpha, \tag{4}$$

which strengthens (3). In general, majorization is a strong form of variability ordering. For example, $a \prec b$ implies that $\sum_i \phi(a_i) \leq \sum_i \phi(b_i)$ for any convex function ϕ. See Marshall and Olkin [13] for further properties and various applications of majorization.

This note presents some majorization results that refine previous work. As a consequence, we prove Milbrodt's strengthening of Hájek's conjecture. Our main results are summarized as follows.

Theorem 1. *Given the drawing probabilities α, let $\pi^{\mathrm{R}}(n)$ (resp., $\pi^{\mathrm{S}}(n)$) denote the first-order inclusion probabilities for rejective sampling (resp., successive sampling) with sample size $n \leq N$.*

Define the "inclusion probabilities per draw" as $p^{\mathrm{R}}(n) \equiv n^{-1}\pi^{\mathrm{R}}(n)$ *and* $p^{\mathrm{S}}(n) \equiv n^{-1}\pi^{\mathrm{S}}(n)$. *Then we have*

$$(N^{-1}, \ldots, N^{-1}) \equiv p^{\mathrm{R}}(N) \prec \cdots \prec p^{\mathrm{R}}(n) \prec \cdots \prec p^{\mathrm{R}}(1) \equiv \alpha, \tag{5}$$

$$(N^{-1}, \ldots, N^{-1}) \equiv p^{\mathrm{S}}(N) \prec \cdots \prec p^{\mathrm{S}}(n) \prec \cdots \prec p^{\mathrm{S}}(1) \equiv \alpha. \tag{6}$$

Moreover,

$$\pi^{\mathrm{R}}(n) \prec \pi^{\mathrm{S}}(n). \tag{7}$$

The ordering chains (5) and (6) are intuitively appealing. Given a set of drawing probabilities, larger sample sizes lead to inclusion probabilities that are more uniform for either rejective sampling or successive sampling. Moreover, (7) says that with the same sample size, the inclusion probabilities are more uniform for rejective sampling than for successive sampling. It is easy to see that (5)–(7) together imply Milbrodt's [14] conjecture, that is, (1) and (2).

We prove (5) and (7) in Section 2 using a combination of analytic and probabilistic techniques. A key tool in resolving (7) is the likelihood ratio order between multivariate densities [9]. A proof of (6), which slightly extends that of (4), is included for completeness.

The Shannon entropy is sometimes used to measure how uniform a distribution is. It is defined as $H(p) = -\sum_{i=1}^{N} p_i \log p_i$ for a probability vector $p = (p_1, \ldots, p_N)$. By convention $0 \log 0 = 0$. It is well known that $p \prec q$ implies $H(q) \le H(p)$. See Cover and Thomas [4], Chapter 2, for further properties of this fundamental quantity. We note the following direct consequence of Theorem 1.

Corollary 1. *In the setting of Theorem* 1,

$$\begin{aligned}
\log N &\equiv H(p^{\mathrm{R}}(N)) \ge \cdots \ge H(p^{\mathrm{R}}(n)) \ge \cdots \ge H(p^{\mathrm{R}}(1)) \equiv H(\alpha),\\
\log N &\equiv H(p^{\mathrm{S}}(N)) \ge \cdots \ge H(p^{\mathrm{S}}(n)) \ge \cdots \ge H(p^{\mathrm{S}}(1)) \equiv H(\alpha),\\
H(p^{\mathrm{R}}(n)) &\ge H(p^{\mathrm{S}}(n)).
\end{aligned}$$

Inequalities are also obtained in terms of the Kullback–Leibler divergence, which is defined as

$$D(p\|q) = \sum_{i=1}^{N} p_i \log \frac{p_i}{q_i}$$

for two probability vectors $p = (p_1, \ldots, p_N)$ and $q = (q_1, \ldots, q_N)$. By convention $x \log(x/0) = \infty$ for $x > 0$ and $0 \log(0/x) = 0$ for $x \ge 0$. A basic property is $D(p\|q) > 0$ unless $p = q$. We shall use $D(p\|q)$ purely as a discrepancy measure between probability vectors without referring to its information-theoretic significance.

Theorem 2. *In the setting of Theorem* 1, *let* $1 \le l < m < n \le N$. *Then we have*

$$D(p^{\mathrm{R}}(l)\|p^{\mathrm{R}}(n)) \ge D(p^{\mathrm{R}}(l)\|p^{\mathrm{R}}(m)) + D(p^{\mathrm{R}}(m)\|p^{\mathrm{R}}(n)), \tag{8}$$

$$D(p^{\mathrm{R}}(n)\|p^{\mathrm{R}}(l)) \geq D(p^{\mathrm{R}}(m)\|p^{\mathrm{R}}(l)) + D(p^{\mathrm{R}}(n)\|p^{\mathrm{R}}(m)), \tag{9}$$

$$D(p^{\mathrm{S}}(n)\|\alpha) \geq D(p^{\mathrm{S}}(n)\|p^{\mathrm{S}}(m)) + D(p^{\mathrm{S}}(m)\|\alpha), \tag{10}$$

$$D(p^{\mathrm{R}}(n)\|\alpha) \geq D(p^{\mathrm{R}}(n)\|p^{\mathrm{S}}(n)) + D(p^{\mathrm{S}}(n)\|\alpha). \tag{11}$$

A number of results can be deduced from these (reverse) triangle inequalities. For example, from (8) and (9) we obtain $D(p^{\mathrm{R}}(m+1)\|\alpha) \geq D(p^{\mathrm{R}}(m)\|\alpha)$ and $D(\alpha\|p^{\mathrm{R}}(m+1)) \geq D(\alpha\|p^{\mathrm{R}}(m))$, showing that, for rejective sampling, the larger the sample size, the more distorted the inclusion probabilities become as compared with the drawing probabilities. Similarly, from (10) we obtain $D(p^{\mathrm{S}}(m+1)\|\alpha) \geq D(p^{\mathrm{S}}(m)\|\alpha)$. From (11) we obtain

$$D(p^{\mathrm{R}}(n)\|\alpha) \geq D(p^{\mathrm{S}}(n)\|\alpha). \tag{12}$$

That is, for fixed n, the inclusion probabilities for successive sampling (rather than for rejective sampling) are more proportional to the drawing probabilities. The inequality (12) may be used to compute an upper bound on $D(p^{\mathrm{S}}(n)\|\alpha)$ because, while $p^{\mathrm{R}}(n)$ can be calculated from α efficiently using a recursive formula (see [3]), numerical calculation of $p^{\mathrm{S}}(n)$ is considerably more difficult.

The inequalities in Theorem 2 resemble the reverse triangle inequalities of Yu [27]. Our results here concern the majorization ordering and may be regarded as first-order results; those in Yu [27] use relative log-concavity and are second order. For related entropy and divergence comparison results, see Karlin and Rinott [10], Johnson [8] and Yu [24–26].

The proof of Theorem 2 builds on Theorem 1 and is presented in Section 3.

2. Proof of Theorem 1

Let $e_k(\cdot)$ denote the kth elementary symmetric function, that is,

$$e_k(\beta) = \sum_{1\leq j_1<\cdots<j_k\leq m} \beta_{j_1}\cdots\beta_{j_k}, \qquad \beta \equiv (\beta_1,\ldots,\beta_m).$$

By convention, $e_0(\beta) \equiv 1$ and $e_k(\beta) = 0$ if $k < 0$ or $k > m$. For a rejective sample of size n, the probability that unit i is included can be expressed as

$$\pi_i^{\mathrm{R}}(n) = \frac{\alpha_i e_{n-1}(\alpha_{-i})}{e_n(\alpha)}, \tag{13}$$

where

$$\alpha_{-i} \equiv (\alpha_1,\ldots,\alpha_{i-1},\alpha_{i+1},\ldots,\alpha_N).$$

The notation $\alpha_{-i,-j}$ (leave-two-out) is defined similarly. It is immediate that $\alpha_i \leq \alpha_j$ implies $\pi_i^{\mathrm{R}}(n) \leq \pi_j^{\mathrm{R}}(n)$. Henceforth, we assume $\alpha_1 \geq \cdots \geq \alpha_N > 0$ without loss of generality.

The following Lemma 1 is needed in the proof of (5).

Lemma 1. *Suppose probability vectors $p = (p_1, \dots, p_N)$ and $q = (q_1, \dots, q_N)$ satisfy*

$$p_1 \geq \cdots \geq p_N > 0, \qquad \frac{q_1}{p_1} \geq \cdots \geq \frac{q_N}{p_N}.$$

Then $p \prec q$.

Proof. For $1 \leq k < N$ we have

$$\frac{\sum_{i=1}^k q_i}{\sum_{i=1}^k p_i} \geq \frac{q_k}{p_k} \geq \frac{q_{k+1}}{p_{k+1}},$$

which yields

$$\frac{\sum_{i=1}^k q_i}{\sum_{i=1}^k p_i} \geq \frac{\sum_{i=1}^{k+1} q_i}{\sum_{i=1}^{k+1} p_i} \geq \cdots \geq \frac{\sum_{i=1}^N q_i}{\sum_{i=1}^N p_i} = 1.$$

Hence $p \prec q$ by definition (the conditions imply $q_1 \geq \cdots \geq q_N$). $\square$

Proof of (5). Let $p \equiv p^{\mathrm{R}}(n+1)$ and $q \equiv p^{\mathrm{R}}(n)$. Note that $\sum_{i=1}^N p_i = \sum_{i=1}^N q_i = 1$. Since $\alpha_1 \geq \cdots \geq \alpha_N$, we have $p_1 \geq \cdots \geq p_N$. The desired relation $p \prec q$ would follow from Lemma 1, if we can show that $q_1/p_1 \geq \cdots \geq q_N/p_N$, or, equivalently, $\pi_k^{\mathrm{R}}(n)/\pi_k^{\mathrm{R}}(n+1) \geq \pi_{k+1}^{\mathrm{R}}(n)/\pi_{k+1}^{\mathrm{R}}(n+1)$ for $1 \leq k < N$. The case $N = 2$ is trivial. Otherwise we have

$$\pi_k^{\mathrm{R}}(n) = \frac{\alpha_k e_{n-1}(\alpha_{-k})}{e_n(\alpha)} = \alpha_k \frac{\alpha_{k+1} e_{n-2}(\tilde{\alpha}) + e_{n-1}(\tilde{\alpha})}{e_n(\alpha)}, \qquad \tilde{\alpha} \equiv \alpha_{-k,-(k+1)}.$$

Thus

$$\frac{\pi_k^{\mathrm{R}}(n)}{\pi_k^{\mathrm{R}}(n+1)} = \frac{e_{n+1}(\alpha)}{e_n(\alpha)} f(\alpha_{k+1}), \tag{14}$$

where

$$f(x) = \frac{x e_{n-2}(\tilde{\alpha}) + e_{n-1}(\tilde{\alpha})}{x e_{n-1}(\tilde{\alpha}) + e_n(\tilde{\alpha})}.$$

Similarly

$$\frac{\pi_{k+1}^{\mathrm{R}}(n)}{\pi_{k+1}^{\mathrm{R}}(n+1)} = \frac{e_{n+1}(\alpha)}{e_n(\alpha)} f(\alpha_k). \tag{15}$$

We have

$$f'(x) = \frac{e_{n-2}(\tilde{\alpha}) e_n(\tilde{\alpha}) - e_{n-1}^2(\tilde{\alpha})}{[x e_{n-1}(\tilde{\alpha}) + e_n(\tilde{\alpha})]^2} < 0,$$

where the inequality follows from Newton's inequalities [7], page 52. That is, $f(x)$ decreases in x. Because $\alpha_{k+1} \le \alpha_k$, we deduce the inequality

$$\frac{\pi_k^{\mathrm{R}}(n)}{\pi_k^{\mathrm{R}}(n+1)} \ge \frac{\pi_{k+1}^{\mathrm{R}}(n)}{\pi_{k+1}^{\mathrm{R}}(n+1)}$$

from (14) and (15). □

The proof of (6) slightly extends and simplifies the arguments of Kochar and Korwar [12].

Proof of (6). Let $S_1, S_2, \ldots \in \{1, \ldots, N\}$ be a sequence of draws retained in successive sampling. It is well known that the inclusion probabilities and the drawing probabilities are ordered in the same way, that is,

$$p_1^{\mathrm{S}}(n) \ge \cdots \ge p_N^{\mathrm{S}}(n), \qquad 1 \le n \le N \tag{16}$$

(see [14]). For $1 \le k \le N$ we have

$$\begin{aligned}
&\Pr(S_n \le k) - \Pr(S_{n+1} \le k) \\
&\quad = \Pr(S_n \le k, S_{n+1} > k) - \Pr(S_n > k, S_{n+1} \le k) \\
&\quad = \sum_{k_1 \le k, k_2 > k} \mathbf{E}[\Pr(S_n = k_1, S_{n+1} = k_2 | S_1, \ldots, S_{n-1}) \\
&\qquad\qquad - \Pr(S_n = k_2, S_{n+1} = k_1 | S_1, \ldots, S_{n-1})],
\end{aligned}$$

where the expectation is with respect to $S_1, \ldots, S_{n-1}$. Because α_i decreases in i, it is easy to show that $k_1 < k_2$ implies

$$\Pr(S_n = k_1, S_{n+1} = k_2 | S_1, \ldots, S_{n-1}) \ge \Pr(S_n = k_2, S_{n+1} = k_1 | S_1, \ldots, S_{n-1}).$$

Hence $\Pr(S_n \le k) \ge \Pr(S_{n+1} \le k)$ for all $1 \le n < N$. This is proved by Kochar and Korwar [12] (see their Lemma 3.2) using a slightly more complicated argument. It follows that

$$\begin{aligned}
\sum_{i=1}^k p_i^{\mathrm{S}}(n) &= n^{-1} \sum_{j=1}^n \Pr(S_j \le k) \\
&\ge (n+1)^{-1} \sum_{j=1}^{n+1} \Pr(S_j \le k) \\
&= \sum_{i=1}^k p_i^{\mathrm{S}}(n+1),
\end{aligned}$$

which proves (6) in view of (16). □

To prove (7), we recall the multivariate likelihood ratio order, also known as the total positivity order (Karlin and Rinott [9], Rinott and Scarsini [17], Shaked and Shanthikumar [21], Chapter 6). Consider the product space $\mathcal{X} = \{1, \dots, N\}^n$. For $x = (x_1, \dots, x_n) \in \mathcal{X}$ and $y = (y_1, \dots, y_n) \in \mathcal{X}$, write

$$x \vee y = (\max\{x_1, y_1\}, \dots, \max\{x_n, y_n\}), \qquad x \wedge y = (\min\{x_1, y_1\}, \dots, \min\{x_n, y_n\}).$$

Let f and g be density functions on $\mathcal{X}$. Then f is said to be no smaller than g in the (multivariate) likelihood ratio order, written as $f \geq_{\mathrm{lr}} g$, if

$$f(x)g(y) \leq f(x \vee y)g(x \wedge y), \qquad x, y \in \mathcal{X}.$$

This generalizes the univariate likelihood ratio order, which requires that the ratio of two univariate densities is a monotone function.

A useful property of the likelihood ratio order is that it implies the usual stochastic order. That is, if X and Y are random vectors taking values in $\mathcal{X}$, and $X \geq_{\mathrm{lr}} Y$ (we use the notation $\geq_{\mathrm{lr}}$ with the random variables as well as their densities), then $\mathbf{E}\phi(X) \geq \mathbf{E}\phi(Y)$ for any coordinatewise increasing function ϕ. In particular, each coordinate of X is no smaller than the corresponding coordinate of Y in the usual stochastic order. Further properties of $\geq_{\mathrm{lr}}$ include closure under marginalization; see Karlin and Rinott [9] and Shaked and Shanthikumar [21], Chapter 6.

Proof of (7). Recall that $\pi_1^{\mathrm{R}}(n) \geq \dots \geq \pi_N^{\mathrm{R}}(n)$. By definition, (7) is proved if we can show

$$\sum_{i=1}^{k} \pi_i^{\mathrm{S}}(n) \geq \sum_{i=1}^{k} \pi_i^{\mathrm{R}}(n), \qquad k = 1, \dots, N-1. \tag{17}$$

Let $X \equiv (X_1, \dots, X_n)$ (resp., $Y \equiv (Y_1, \dots, Y_n)$) denote the unit indices arranged in increasing order of a sample of size n obtained by rejective sampling (resp., successive sampling). That is, X and Y take values in $\Omega \equiv \{(x_1, \dots, x_n) \in \mathcal{X}\colon 1 \leq x_1 < \dots < x_n \leq N\}$. Then an unnormalized density of X is

$$f(x) = \alpha_{x_1} \dots \alpha_{x_n}, \qquad x = (x_1, \dots, x_n) \in \Omega,$$

and the density of Y can be written as

$$\begin{aligned} g(y) &= \sum_{\sigma \in \mathrm{Perm}(y)} \alpha_{\sigma_1} \frac{\alpha_{\sigma_2}}{1 - \alpha_{\sigma_1}} \frac{\alpha_{\sigma_3}}{1 - \alpha_{\sigma_1} - \alpha_{\sigma_2}} \cdots \frac{\alpha_{\sigma_n}}{1 - \sum_{j=1}^{n-1} \alpha_{\sigma_j}} \\ &= \sum_{\sigma \in \mathrm{Perm}(y)} \frac{\alpha_{y_1} \cdots \alpha_{y_n}}{(1 - \alpha_{\sigma_1})(1 - \alpha_{\sigma_1} - \alpha_{\sigma_2}) \cdots (1 - \sum_{j=1}^{n-1} \alpha_{\sigma_j})}, \qquad y = (y_1, \dots, y_n) \in \Omega, \end{aligned}$$

where $\sigma = (\sigma_1, \dots, \sigma_n)$ and $\mathrm{Perm}(y)$ denotes the set of vectors obtained by permuting the coordinates of y. Note that, for $x, y \in \Omega$ we have $x \vee y \in \Omega$ and $x \wedge y \in \Omega$. Moreover, for

$x, y \in \Omega$,

$$\begin{aligned} f(x)g(y) &= \sum_{\sigma \in \mathrm{Perm}(y)} \frac{\alpha_{x_1} \cdots \alpha_{x_n} \alpha_{y_1} \cdots \alpha_{y_n}}{(1-\alpha_{\sigma_1})(1-\alpha_{\sigma_1}-\alpha_{\sigma_2}) \cdots (1-\sum_{j=1}^{n-1} \alpha_{\sigma_j})} \\ &\le \sum_{\sigma \in \mathrm{Perm}(x \wedge y)} \frac{\alpha_{x_1} \cdots \alpha_{x_n} \alpha_{y_1} \cdots \alpha_{y_n}}{(1-\alpha_{\sigma_1})(1-\alpha_{\sigma_1}-\alpha_{\sigma_2}) \cdots (1-\sum_{j=1}^{n-1} \alpha_{\sigma_j})} \\ &= f(x \vee y) g(x \wedge y), \end{aligned}$$

where the inequality holds because α_i decreases in i and, under an obvious bijection, each element in $\mathrm{Perm}(y)$ is at least as large as its counterpart in $\mathrm{Perm}(x \wedge y)$. Thus $X \ge_{\mathrm{lr}} Y$. It follows that

$$\Pr(X_j \le k) \le \Pr(Y_j \le k), \qquad j = 1, \dots, n, k = 1, \dots, N.$$

That is, X_j is no smaller than Y_j in the usual stochastic order. We have, for $1 \le k \le N$,

$$\begin{aligned} \sum_{i=1}^{k} \pi_i^{\mathrm{R}}(n) &= \sum_{i=1}^{k} \sum_{j=1}^{n} \Pr(X_j = i) \\ &= \sum_{j=1}^{n} \Pr(X_j \le k) \\ &\le \sum_{j=1}^{n} \Pr(Y_j \le k) \\ &= \sum_{i=1}^{k} \sum_{j=1}^{n} \Pr(Y_j = i) \\ &= \sum_{i=1}^{k} \pi_i^{\mathrm{S}}(n). \end{aligned}$$

Thus (17) holds, and the proof is complete. □

3. Proof of Theorem 2

The following Lemma 2 is key to the proof of Theorem 2.

Lemma 2. *Let $p = (p_1, \dots, p_N)$, $q = (q_1, \dots, q_N)$ and $r = (r_1, \dots, r_N)$ be probability vectors with all positive coordinates. If either* (a) $q \prec p$, $p_1 \ge \cdots \ge p_N$, *and* $q_1/r_1 \ge \cdots \ge q_N/r_N$, *or* (b) $p \prec q$, $q_1 \ge \cdots \ge q_N$, *and* $q_1/r_1 \le \cdots \le q_N/r_N$, *then*

$$D(p\|r) \ge D(p\|q) + D(q\|r).$$

Proof. Let us assume (a). Case (b) is similar. We have

$$
\begin{aligned}
D(p\|r) - D(p\|q) - D(q\|r) &= \sum_{i=1}^{N} (p_i - q_i) \log \frac{q_i}{r_i} \\
&= \sum_{i=1}^{N-1} \left(\sum_{j=1}^{i} p_j - \sum_{j=1}^{i} q_j \right) \left(\log \frac{q_i}{r_i} - \log \frac{q_{i+1}}{r_{i+1}} \right) \qquad (18) \\
&\geq 0,
\end{aligned}
$$

where the first equality follows from the definition of the Kullback–Leibler divergence, the second equality holds by summation by parts, and the inequality holds because q_i/r_i decreases in i and $q \prec p$, and hence both parentheses in (18) are non-negative. □

As in Section 2, in the proofs of (8)–(11) we assume $\alpha_1 \geq \cdots \geq \alpha_N > 0$.

Proof of (8) and (9). Let $p \equiv p^{\mathrm{R}}(l), q \equiv p^{\mathrm{R}}(m), r \equiv p^{\mathrm{R}}(n)$. Then $p_1 \geq \cdots \geq p_N$. Since $l < m$ we have $q \prec p$ by (5). From the proof of (5) we know that $q_1/r_1 \geq \cdots \geq q_N/r_N$. Thus (8) follows from Lemma 2, Case (a). The proof of (9) is similar. □

To prove (10) and (11) we need the following result.

Proposition 1. *The ratio* $p_i^{\mathrm{S}}(n)/\alpha_i, i = 1, \ldots, N$, *increases in* i *for each* $n \leq N$.

Proof. Let $\pi_{i,k}$ denote the probability that the kth distinct draw in successive sampling is unit i. Then $p_i^{\mathrm{S}}(n) = n^{-1} \sum_{k=0}^{n-1} \pi_{i,k+1}$. It suffices to show that $\pi_{i,k+1}/\alpha_i$ increases in i for each k. Let us assume $k \geq 1$ and define the index set

$$\Omega(i) = \{(j_1, \ldots, j_k) \colon 1 \leq j_l \leq N, j_l \neq i, 1 \leq l \leq k, \text{ and } j_l \text{ are distinct}\}.$$

Then we have

$$\frac{\pi_{i,k+1}}{\alpha_i} = \sum_{(j_1,\ldots,j_k)\in\Omega(i)} \alpha_{j_1} \frac{\alpha_{j_2}}{1-\alpha_{j_1}} \cdots \frac{\alpha_{j_k}}{1-\sum_{l=1}^{k-1}\alpha_{j_l}} \left(\frac{1}{1-\sum_{l=1}^{k}\alpha_{j_l}} \right). \qquad (19)$$

The summand is a decreasing function in $(j_1, \ldots, j_k)$, since α_j decreases in j. Consider a mapping $\Omega(i) \to \Omega(i+1)$ that sends $(j_1, \ldots, j_k) \in \Omega(i)$ to $(j_1^*, \ldots, j_k^*) \in \Omega(i+1)$ as follows. For $l = 1, \ldots, k$, if $j_l \neq i+1$, let $j_l^* = j_l$; otherwise let $j_l^* = i$. It is easy to see that this mapping is well defined and is a bijection. Note that $j_l^* \leq j_l$. Hence the right-hand side of (19) increases if we replace the summation index $\Omega(i)$ by $\Omega(i+1)$. That is, $\pi_{i,k+1}/\alpha_i$ increases in i, as required. □

Proof of (10) and (11). Let $p \equiv p^{\mathrm{R}}(n), q \equiv p^{\mathrm{S}}(n)$ and $r \equiv \alpha$. By Proposition 1, $q_1/r_1 \leq \cdots \leq q_N/r_N$. By (7) we have $p \prec q$. Thus (11) follows from Lemma 2, Case (b). The proof of (10) is similar. □

References

[1] Arratia, R., Goldstein, L. and Langholz, B. (2005). Local central limit theorems, the high-order correlations of rejective sampling and logistic likelihood asymptotics. *Ann. Statist.* **33** 871–914. MR2163162

[2] Berger, Y.G. (1998). Rate of convergence to normal distribution for the Horvitz-Thompson estimator. *J. Statist. Plann. Inference* **67** 209–226. MR1624693

[3] Chen, X.H., Dempster, A.P. and Liu, J.S. (1994). Weighted finite population sampling to maximize entropy. *Biometrika* **81** 457–469. MR1311090

[4] Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons]. MR2239987

[5] Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Statist.* **35** 1491–1523. MR0178555

[6] Hájek, J. (1981). *Sampling from a Finite Population*. *Statistics*: *Textbooks and Monographs* **37**. New York: Dekker Inc. Edited by Václav Dupač, With a foreword by P. K. Sen. MR0627744

[7] Hardy, G.H., Littlewood, J.E. and Pólya, G. (1964). *Inequalities*. Cambridge, UK: Cambridge Univ. Press.

[8] Johnson, O. (2007). Log-concavity and the maximum entropy property of the Poisson distribution. *Stochastic Process. Appl.* **117** 791–802. MR2327839

[9] Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.* **10** 467–498. MR0599685

[10] Karlin, S. and Rinott, Y. (1981). Entropy inequalities for classes of probability distributions. I. The univariate case. *Adv. in Appl. Probab.* **13** 93–112. MR0595889

[11] Kaufman, G.M. (1996). Successive sampling and software reliability. *J. Statist. Plann. Inference* **49** 343–369. MR1381164

[12] Kochar, S.C. and Korwar, R. (2001). On random sampling without replacement from a finite population. *Ann. Inst. Statist. Math.* **53** 631–646. MR1868896

[13] Marshall, A.W. and Olkin, I. (1979). *Inequalities*: *Theory of Majorization and Its Applications*. *Mathematics in Science and Engineering* **143**. New York: Academic Press [Harcourt Brace Jovanovich Publishers]. MR0552278

[14] Milbrodt, H. (1992). Comparing inclusion probabilities and drawing probabilities for rejective sampling and successive sampling. *Statist. Probab. Lett.* **14** 243–246. MR1173625

[15] Qualité, L. (2008). A comparison of conditional Poisson sampling versus unequal probability sampling with replacement. *J. Statist. Plann. Inference* **138** 1428–1432. MR2388021

[16] Rao, T.J., Sengupta, S. and Sinha, B.K. (1991). Some order relations between selection and inclusion probabilities for PPSWOR sampling scheme. *Metrika* **38** 335–343. MR1144976

[17] Rinott, Y. and Scarsini, M. (2006). Total positivity order and the normal distribution. *J. Multivariate Anal.* **97** 1251–1261. MR2276157

[18] Rosén, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement. I, II. *Ann. Math. Statist.* **43** 373–397; ibid. 43 (1972), 748–776. MR0321223

[19] Rosén, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement. II. *Ann. Math. Statist.* **43** 748–776. MR0321223

[20] Rosén, B. (2000). On inclusion probabilities for order πps sampling. *J. Statist. Plann. Inference* **90** 117–143. MR1791585

[21] Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*. *Springer Series in Statistics*. New York: Springer. MR2265633

[22] Stern, H. and Cover, T.M. (1989). Maximum entropy and the lottery. *J. Amer. Statist. Assoc.* **84** 980–985. MR1134487

[23] Traat, I., Bondesson, L. and Meister, K. (2004). Sampling design and sample selection through distribution theory. *J. Statist. Plann. Inference* **123** 395–413. MR2062989
[24] Yu, Y. (2008). Letter to the editor: On an inequality of Karlin and Rinott concerning weighted sums of i.i.d. random variables [MR0595889]. *Adv. in Appl. Probab.* **40** 1223–1226. MR2488539
[25] Yu, Y. (2008). On the maximum entropy properties of the binomial distribution. *IEEE Trans. Inform. Theory* **54** 3351–3353. MR2450793
[26] Yu, Y. (2009). On the entropy of compound distributions on nonnegative integers. *IEEE Trans. Inform. Theory* **55** 3645–3650. MR2598065
[27] Yu, Y. (2010). Relative log-concavity and a pair of triangle inequalities. *Bernoulli* **16** 459–470. MR2668910

Received June 2010